

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 00:49:37 ; Search time 2859.69 Seconds  
(without alignments)  
13738.812 Million cell updates/sec

Title: US-09-768-781-1

Perfect score: 1350

Sequence: 1 atggcagagtttgaat.....caagcagaagtgtgtctga 1350

Scoring table:

IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hgt: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sta: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sta: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	747.2	55.3	37650	9	HSU131B10
C 2	620.8	46.0	168918	2	AC116256 Human DNA s
C 3	615.4	45.6	171014	10	AL671915 Rattus no
4	441.2	32.7	1609	9	AK098608 Mouse DNA
5	301.8	22.4	2311	10	AF155511 Homo sapi
6	289.6	21.5	5096	6	AX332514 Mus muscu
7	289.6	21.5	5096	9	HSXKMTF
8	282.8	20.9	5215	6	AX405691 Sequence
C 9	276.4	20.5	113688	9	AC005301 Homo sapi
C 10	276.4	20.5	124823	9	AC007064 Homo sapi
C 11	249.8	18.5	168918	2	AC116256 Rattus no
C 12	218	16.1	160200	10	AC091605 Mus muscu
C 13	218	16.1	200698	2	AL672060 Mus muscu
C 14	215.6	16.0	166436	9	AC016752 Homo sapi
15	214	15.9	161879	9	AC017005 Homo sapi
16	214	15.9	182083	9	AC007965 Homo sapi
17	209.2	15.5	37566	9	AC068541 Homo sapi
C 18	209.2	15.5	101912	9	AC000100 Homo sapi
C 19	209.2	15.5	120951	9	AC073649 Homo sapi
C 20	209.2	15.5	145383	9	AC010682 Homo sapi
C 21	209.2	15.5	149794	2	AC007273 Homo sapi
C 22	209.2	15.5	174082	9	AC007379 Homo sapi
23	194.4	14.4	175531	9	HSXKSRPM
24	178.2	13.2	163495	9	AC024183 Homo sapi
C 25	178.2	13.2	164891	2	AC022848 Homo sapi
C 26	178.2	13.2	199174	9	AC007742 Homo sapi
27	171.4	12.7	1588	9	HSXF000997
C 28	164	12.1	91640	2	AC123295 Rattus no
29	160.2	11.9	826	6	AX079777 Sequence
30	103.2	7.6	321	11	G03706 human STS W
C 31	87.6	6.5	164891	2	AC022848 Homo sapi
C 32	60.4	4.5	64450	9	AC073893 Homo sapi
C 33	57.8	4.3	138674	2	AC112084 Rattus no
34	51.6	3.8	125020	9	AF429315 Homo sapi
35	49.8	3.7	188048	10	AL669974 Mouse DNA
C 36	46.8	3.5	68873	2	AC101531 Mus muscu
C 37	46.8	3.5	234491	2	AL671118 Mus muscu
38	45.8	3.4	212374	2	AC022773 Mus muscu
C 39	45.6	3.4	144794	9	AC107055 Homo sapi
40	45.4	3.4	7218	6	I66494 Sequence 14
41	44	3.3	152434	2	AC016262 Homo sapi
C 42	43.4	3.2	1694	9	AF226053 Homo sapi
43	43.2	3.2	12229	3	AE001377 Plasmodi
44	42.8	3.2	13427	6	AX346828 Sequence
45	42.2	3.1	104455	2	AC098390 Rattus no

# ALIGNMENTS

RESULT 1  
HSU131B10  
LOCUS  
DEFINITION  
Human DNA sequence from cosmid U131B10, between markers DXS366 and DXS87 on chromosome X contains XK membrane transport protein, ESTs and STS.  
37650 bp DNA linear PRI 23-NOV-1999

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

273417  
GI:1322397  
X; XK membrane transport protein.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 37650)  
Bodsworth, S.

Pred. No. is the number of results predicted by chance to have a





## TITLE

Direct Submission  
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:19718450.

## COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTLM

Center clone name: CH230-139K19

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 104166 bases at least Q40

Consensus quality: 113489 bases at least Q30

Consensus quality: 119214 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 74 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

'runs of N', but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1568: contig of 1568 bp in length

1569 1668: gap of unknown length

1669 2687: contig of 1019 bp in length

2688 2787: gap of unknown length

2788 4137: contig of 1350 bp in length

4138 4237: gap of unknown length

4238 5268: contig of 1031 bp in length

5269 5368: gap of unknown length

5369 6665: contig of 1297 bp in length

6666 6765: gap of unknown length

6766 8226: contig of 1461 bp in length

8227 9439: gap of unknown length

9440 9549: contig of 1123 bp in length

9550 10623: contig of 1074 bp in length

10624 10723: gap of unknown length

10724 11766: contig of 1043 bp in length

11767 11866: gap of unknown length

11867 12876: contig of 1010 bp in length

12877 12976: gap of unknown length

12977 14300: contig of 1324 bp in length

14301 14400: gap of unknown length

14401 15614: contig of 1214 bp in length

15615 15714: gap of unknown length

15715 17224: contig of 2010 bp in length

17225 17824: gap of unknown length

17825 19055: contig of 1231 bp in length

19056 19155: gap of unknown length

19156 20675: contig of 1520 bp in length

20676 20775: gap of unknown length

20776 22570: contig of 1795 bp in length

22571 22670: gap of unknown length

22671 24166: contig of 1496 bp in length

24167 24266: gap of unknown length

24267 25499: contig of 1233 bp in length

25500 25599: gap of unknown length

25600 27074: contig of 1475 bp in length

27075 27174: gap of unknown length

27175 28475: contig of 1301 bp in length

28476 28575: gap of unknown length

28576 29878: contig of 1303 bp in length

29879 29978: gap of unknown length

29979

31798: contig of 1820 bp in length

31799 gap of unknown length

31800 contig of 1558 bp in length

31801 gap of unknown length

31802 contig of 1197 bp in length

31803 gap of unknown length

31804 contig of 2224 bp in length

31805 gap of unknown length

31806 contig of 1782 bp in length

31807 gap of unknown length

31808 contig of 1544 bp in length

31809 gap of unknown length

31810 contig of 1176 bp in length

31811 gap of unknown length

31812 contig of 2046 bp in length

31813 gap of unknown length

31814 contig of 1226 bp in length

31815 gap of unknown length

31816 contig of 1558 bp in length

31817 gap of unknown length

31818 contig of 2327 bp in length

31819 gap of unknown length

31820 contig of 1904 bp in length

31821 gap of unknown length

31822 contig of 1978 bp in length

31823 gap of unknown length

31824 contig of 1202 bp in length

31825 gap of unknown length

31826 contig of 1447 bp in length

31827 gap of unknown length

31828 contig of 1135 bp in length

31829 gap of unknown length

31830 contig of 1901 bp in length

31831 gap of unknown length

31832 contig of 2048 bp in length

31833 gap of unknown length

31834 contig of 2987 bp in length

31835 gap of unknown length

31836 contig of 1711 bp in length

31837 gap of unknown length

31838 contig of 2792 bp in length

31839 gap of unknown length

31840 contig of 2331 bp in length

31841 gap of unknown length

31842 contig of 1722 bp in length

31843 gap of unknown length

31844 contig of 1346 bp in length

31845 gap of unknown length

31846 contig of 2527 bp in length

31847 gap of unknown length

31848 contig of 1733 bp in length

31849 gap of unknown length

31850 contig of 1429 bp in length

31851 gap of unknown length

31852 contig of 2576 bp in length

31853

Query Match 46.0%; Score 620.8; DB 2; Length 168918;

Best Local Similarity 88.4%; Pred. No. 2.8e-168;

Matches 673; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 590 TTCCCTGGGTAGAGTTTGGCTAATGGTATTTTCCCTGGTATCTGTCCCTATGGGCCA 649

DB 140555 TTCCCTATTTTGCACCTGTGTTAATGTTTTTCCCTGATATCTGTCCCTATGGGCTA 140496

QY 650 CCCTTGGCAATATGTTGGCTATCCAGATCRAAGTACGATCAAGATTCGCTTGGGC 709  
 DB 140495 CCCTCTGCAATATGTTGGCTATCCAGATCAAGTATGATGATCAAGATTCGCTTGGTC 140436  
 QY 710 CACTAGAAAGTCTCTGATCAACCATCTGGCGGACATTTGGAGATCACTTCCCGCTCTCTGA 769  
 DB 140435 CAGTGAAGTCTCTGATCAACCATCTGGAGGACATTTGGAGATCACTTCCCGCTCTATGA 140376  
 QY 770 TTCTGGTCTCTCTCAGCACTTTGAAATTAAGAGCTGTGCTTCTAGTGTCTCAACT 829  
 DB 140375 TTCTGGTCTCTCTCAGCACTTTGAAATTAAGAGCTGTGCTTCTAGTGTCTCAACT 140316  
 QY 830 TCCTGATCATCTCTTTGAGCCCTGGATTAAAGTTCTGGAGAGGCTGCCAGATGCCCA 889  
 DB 140315 TCTTGATCATCTCTTTGAGCCCTGGATTAAAGTTCTGGAGAGGCTGCCAGATGCCCA 140256  
 QY 890 ATAACTTTGAGAAAACCTTTACCGCGGTGCGCACTCTGGTGGTCTGCTGATTTTCAGTCACCA 949  
 DB 140255 ACAATTTGAGAAAATTTTCAGCCGAGTTGGACCGCTGGTGTACTGATTTCTGTACCA 140196  
 QY 950 TCCTCTATGCTGGCAATCACTTCTCTGCTGGTCACTTTGCACTTTCAGTTGAGGTGGCAGCA 1009  
 DB 140195 TCCTCTATGCTGGCAATCACTTCTCTGCTGGTCACTTTGCACTTTCAGTTGAGGTGGCAGCA 140136  
 QY 1010 GAGATCTCTCGACAAAGGCGAGACTGGGACATATGGGCTGCACTATAGTGTAGGT 1069  
 DB 140135 GAGACCTCTGTTGACAAAGGTCAGAACTGGGACATATGGGACTGCACCTACAGTGTAGGT 140076  
 QY 1070 TGGTAGAAGATGTCATCATGCTCTTGGTTTAAAGTTCTTTGGAGTGAAGTGTACTGA 1129  
 DB 140075 TGGTAGAAGATGTCATCATGCTCTTGGTTTAAAGTTCTTTGGAGTGAAGTGTACTGA 140016  
 QY 1130 ATTACTGTCATCTCTGATTCGCTTGCAGTCACTATTTGCTTATCTGATTTCCATTTGACT 1189  
 DB 140015 ATTACTGTCATCTCTGATTCGCTTGCAGTCACTATTTGCTTATCTGATTTCCATTTGACT 139956  
 QY 1190 TCATGCTCTCTCTTCCAGTACTTGCATCCATTTGGCTGCTACTCTTACCCTATATGTAG 1249  
 DB 139955 TCATGCTCTCTCTTCCAGTACTTGCATCCATTTGGCTGCTACTCTTACCCTATATGTAG 139896  
 QY 1250 TAGACTACCTCCATTTGCTGCTGTCCAGCAGCACCCTCGGACAGGCTTGGAGAACTCAG 1309  
 DB 139895 TAGACTACCTCCATTTGCTGCTGTCCAGCAGCACCCTCGGACAGGCTTGGAGAACTCAG 139836  
 QY 1310 AGCCACCTTTGAGACTGAAGCAAGGCAAGTGTGTCTGA 1350  
 DB 139835 AGACATCTGTGAGCTGAGCCTGAGACAAAGGCAAGTGTGTCTGA 139795

RESULT 3  
 AL671915/c  
 LOCUS Mouse DNA sequence from clone Rp23-182N4 on chromosome X, complete sequence.  
 DEFINITION AL671915  
 ACCSSION AL671915  
 VERSION AL671915.8 GI:21537500  
 KEYWORDS HTG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 171014)  
 Heath, P.  
 Direct Submission  
 Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphries@sanger.ac.uk  
 On Jun 21, 2002 this sequence version replaced gi:21213424.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Rp23-182N4 is from the RpCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6.

# FEATURES

Location/Qualifiers  
 1..171014  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="Rp23-182N4"  
 /clone\_lib="RpCI-23"

BASE COUNT 50391 a 34981 c 35121 g 50521 t  
 ORIGIN

Query Match 45.6%; Score 615.4; DB 10; Length 171014;  
 Best Local Similarity 88.0%; Pred. No. 1e-166;  
 Matches 670; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
 QY 590 TTCCCTGGGTAGAGTTGTCTAATGTAATTTCCCTGGTATCTGTCACTATGGGGCA 649  
 DB 115391 TTCTTATTTTGCAGCTGTGTTAATGGCTTTTCCCTGATATCTGTCACTACGGGGCTA 115332  
 QY 650 CCCTTTGCATATGTTGGCTATCCAGATCAAGTACGATCAAGATTCGCTTGGGC 709  
 DB 115331 CCCTGTGCAATATGTTGGCTATCCAGATCAAGTATGATGATCAAGATTTAGACTGGGGC 115272  
 QY 710 CACTAGAAAGTCTCTGATCAACCATCTGGGACATTTGGAGATCACTTCCCGCTCTCTGA 769  
 DB 115271 CACTGGAAGTCCCTTGTATCACCGTCTGGAGGACATTTGGAGATCACTTCCCGCTCTCTGA 115212  
 QY 770 TTCTGGTCTCTTCTCAGCCACTTTGAAATTTGAAGGCTGTGCCCTTCTAGTGTCTCACT 829  
 DB 115211 TTCTGGTCTCTTCTCAGCCACTTTGAAATTTGAAGGCTGTGCCCTTCTAGTGTCTCACT 115152  
 QY 830 TCCTGATCATCTCTTTGAGCCCTGGATTAAAGTTCTGGAGAGGCTGGTCCAGATGCCCA 889  
 DB 115151 TCTTGATCATCTCTTTGAGCCCTGGATTAAAGTTCTGGAGAGGCTGGTCCAGATGCCCA 115092  
 QY 890 ATAACTTTGAGAAAACCTTTACCGCGGTGCGCACTCTGGTGGTCTGCTGATTTTCAGTCACCA 949  
 DB 115091 ACAATTTGAGAAAATTTTCAGCCGAGTTGGACCGCTGGTGTACTGATTTCTGTACCA 115032  
 QY 950 TCCTCTATGCTGGCATCAACTTCTCTGCTGGTCACTTTTGGAGTGAAGTGTGGCAGACA 1009  
 DB 115031 TTCTCTATGCTGGCATCAACTTCTCTGCTGGTCACTTTTGGAGTGAAGTGTGGCAGACA 114972  
 QY 1010 GAGATCTCTCGACAAAGGCGAGACTGGGACATATGGGCTGCACTATAGTGTAGGT 1069  
 DB 114971 GAGACCTTGTGACAAAGGTCAGAACTGGGACATATGGGACTGCACCTATAGTGTAGGC 114912  
 QY 1070 TGGTAGAAGATGTCATCATGCTCTTGGTTTAAAGTTCTTTGGAGTGAAGTGTACTGA 1129  
 DB 114911 TGGTAGAAGATGTCATCATGCTCTTGGTTTAAAGTTCTTTGGAGTGAAGTGTACTGA 114852  
 QY 1130 ATTACTGTCATCTCTGATTCGCTTGCAGTCACTATTTGCTTATCTGATTTCCATTTGACT 1189  
 DB 114851 ATTACTGTCATCTCTGATTCGCTTGCAGTCACTATTTGCTTATCTGATTTCCATTTGACT 114792  
 QY 1190 TCATGCTCTCTCTTCCAGTACTTGCATCCATTTGGCTGCTACTCTTCCACCCATATGTAG 1249

||||| 114731 TCATGCTTCTTCTTCCAGTACTTGCACCCCTGCGCTCACTTCCACCAATACGTAG 114732  
 1250 TAGACTTACTCCATTGTGCTGTCTCACCAGCACCCCTCGACAGAGGTTGAGAACTCAG 1309  
 114731 TAGACTTACTCCATTGTGCTGTCTCACCAGCACCCCTCGGAAAGGTTGAGAACTCAG 114672  
 1310 AGCCACCCCTTGAGACTGAAGCAAGCGCAAGTGTGTCTGA 1350  
 114671 AGACATCCTGTGAAGCTGACACACACAAAGTATTGTCTGA 114631

AK098608 1609 bp mRNA linear PRI 15-JUL-2002  
 Homo sapiens CDNA FLJ25742 fis, clone TST05965.  
 AK098608  
 AK098608.1 GI:21758666  
 oligo capping: fis (full insert sequence).  
 Homo sapiens testis cDNA to mRNA, clone\_lib:TST clone:TST05965.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,  
 Arita, M., Mueasino, K., Yuuki, H., Hara, H., Suzuki, Y., Hata, H.,  
 Nakagawa, K., Mizuno, S., Morinaga, M., Kawamura, M., Sugiyama, T.,  
 Irie, R., Otsuki, T., Sato, H., Nishikawa, T., Sugiyama, A.,  
 Kawakami, B., Nagai, K., Isogai, T. and Sugano, S.  
 NEDO human cDNA sequencing project.  
 Unpublished  
 2 (bases 1 to 1609)  
 Sugano, S. and Suzuki, Y.  
 Direct Submission  
 Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Laboratory of Genome Structure, Human Genome  
 Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
 (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
 Fax:81-3-5449-5416)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; CDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); CDNA library  
 construction and 5'-end one pass sequencing: Institute of Medical  
 Science, University of Tokyo, Laboratory of Genome Structure, Human  
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for  
 full insert sequencing: RAB and Helix Research Institute.  
 Location/Qualifiers  
 1..1609  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="TST05965"  
 /tissue\_type="testis"  
 /clone\_libs="TST"  
 /note="cloning vector: pME18SFL3"  
 93..1472  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAC05352.1"  
 /db\_xref="GI:21758667"  
 /translation="METVFENEDESTGVSSEKEIVLQRLHLSPFPISTVLYL  
 CEVAFGLYMEIYKANDTFWMSFTIIVGALDQIILMFNKLRRNKAALFEW  
 HTLLGSPVRLCHTIRNYHKMLKNLQKEETQVSTIKRNTMLREIAFSTRDNPMQO  
 KAFKYSVIOALPGSVPOLIQMYLSKIREWPNRALLMTFSLISVTYTGARCNILA  
 TQISNDTDTKLPLIEFFCVMMRLEVISRVVTALFIASLKULSPVLVLIYFVSI  
 LAPLEWFKSGAHLPGKKNENSMVGMFLITLLYAAINFCSKSAVKQLSDDDKI  
 IDGRGWRGHRIUHYSGFLENVIMILVFRFGKTLNCCDSLSIAVLIIISYLLATGFI  
 MLLFYQILYPMQSGKVLPORTENPEAPYVYVNIETKKNKNKQLRNYCNSNRVGYH  
 SIRKSWTCS"

486 a 290 c 336 g 497 t

BASE COUNT  
 RIGIN

Query Match 32.7%; Score 441.2; DB 9; Length 1609;  
 Best Local Similarity 60.7%; Pred. No. 1.8e-116;

Matches	747;	Conservative	0;	Mismatches	468;	Indels	15;	Gaps	1
QY	1	ATGGCAGAGTTTTATGAAATTCCTGAGGAGCCAAATGTGGATCGGTTTCATCTCTGGAG	60						
Db	93	ATGGACACAGTGTTCGAAGAGATGGATGAAGAAAGCACAGGAGGATTTTCATCTTCGAAA	152						
QY	61	GAAGATGTCATCGGTGGAGCAACCCCGATTTACTTTTCCATTTAGCATCCTTTTCTCC	120						
Db	153	GAAGAATAATGTCCTGGCCAGAGACTCCATCTAAGCTTTCCTTTTATAGCATTTATCTCTCA	212						
QY	121	ACCTTTTGTACTGGGAGGCTGCATCTGCTTTGTACATGTTAGAACTATCGAAAG	180						
Db	213	ACTGTTCTCTACTGTGTGAGGTGGCTTTGGTTTATACATGTTTGAATTTATCGAAA	272						
QY	181	AATAGTGAACCTTACCGGATGACATACACCTTTTCTTTTATGTTTTCATCAATTATG	240						
Db	273	GCTAATGACACATTTCTGGATGTCATTTACCATCAGCTTTATTTATTTGGGGCAATTTTG	332						
QY	241	GTCAGTTGACCTCATTTTGTGTCACAGAGATCTAGCCAAAGNATAACCGCTATCATTA	300						
Db	333	GATCAAAATATCTCTGATGTTTTTCAACAAAGACTTGAGGAGNAATAAGGCTGCAATTACTT	392						
QY	301	TTTATGCATCTAATCTCTCTGGACCTGTTTATCAGATGTTTGAGGCGCATGATTAAGTAC	360						
Db	393	TTTTGGCAATCTCTTTTATAGACCTATTGTGAGGTGTTGACACCATTAGAATATAC	452						
QY	361	CTCACACTGTGGAAGAAAGAGGACGAGGAGGCCCTTATGTCAGCTCACCCGAAAGAG	420						
Db	453	CACAAATGGTTGAAAATCTTAACAGGAGAGGAGAGACTCAAGTTAGCATCACAAAG	512						
QY	421	ATGCTAATAGATGCGGAGGAGGTGCTGTATAGAATGGAGGTGGGCCACTGCATCCGAC	480						
Db	513	AGAAACAC-----GATGCTGGAAGGAGATTTGCAATCTCAATCCGGAT	557						
QY	481	CTGGCTATGCACGCAATGCTACAAACGATGTGTCAGATCCAAGCTTCTCGGGCTCA	540						
Db	558	AATTCATGACAGAGAGGCTTCAAGTACATGTCAGTGATCAGGCTTTCTCGTTCT	617						
QY	541	GTGCCCAGCTGACCTATCAGCTCTATGTGAGCGCTGATCTCTGACAGAGTTTCCCTGGGT	600						
Db	618	GTTCACAAATTAATTTTGCAGATGTATATCAGTCTACTATACGAGAATGGCTTTGAAT	677						
QY	601	AGAGTTGTGCTAATGTGTAATTTTCCCTGGTATCTGTACCTATGGGGCCACCTTTGCAAT	660						
Db	678	AGAGCATTTGCTGATGACATTTTCCCTGTTATCAGTTACTTATGGGGCCATTCGCTGCAAT	737						
QY	661	ATGTTGCTATCCAGATCAAGTACGATGACTACAAGATTCGCCCTTGGSCCATAGAACTC	720						
Db	738	ATACTGGCCATCAGATCAGCAATGATGATCTACCAATTAAGCTACCGCTGATAGAAATC	797						
QY	721	CTCTGCATCACCATCTGGCGGACATTTGGAGATCACTTCCCGCTTCTGATTTCTGGTGCTC	780						
Db	798	TTCTGTGCTGTGATGTGGCGTTTTTTGGAGGTTATCTCACGTGTAGTACTGGCAATTG	857						
QY	781	TTCTCAGCCACTTTTGAATAATGAAGGCTGTGCCCTTCCGTAGTGCTCAACTTCCGTGATC	840						
Db	858	TTCAATGCATCTCTGAAACTGAAGAGGCTACCCCGTTTTGTTTAATCATATATTTTGTATCA	917						
QY	841	CTCTTTGAGCCCTGGATTAAGTTCTCGAGAGTGGTGCCAGATGCCCAATAACATTGAG	900						
Db	918	TTGTTGGCAACGTTGGCTGGAGTTTTTGGAAAAGTGGAGCTCATCTTCTTGGCAACAAGAA	977						
QY	901	AAAACTTTCAGCCGGGTGGGCACTCTGGTGCTCTGATTTTCAGTCACCATCTCTATGCT	960						
Db	978	AATPAATTCGAATATGGTGGGTACAGTACTGATGCTTTTCTTGATCACACTGCTATATGCT	1037						
QY	961	GGCATCAACTCTCTTCTGTGTGCTAGCTTTTGCAGTTTGGAGTTGGGCGACAGAGATCTCGTC	1020						
Db	1038	GCCATCAACTCTCTGCTGTGTCAGCAGTGAACCTGCACTGCTGATGACAGAAATAAT	1097						
QY	1021	GACAAAGGGCAGACTGGGACATATGGGCTCGCATATAGTTGAGGTTGGTAGAGAT	1080						
Db	1098	GACGGGAGACAGAGGTGGGGCCATAGAATCTTACATACAGCTTTCAGTTTTCAGAAAT	1157						

||||| 114731 TCATGCTTCTTCTTCCAGTACTTGCACCCCTGCGCTCACTTCCACCAATACGTAG 114732  
 1250 TAGACTTACTCCATTGTGCTGTCTCACCAGCACCCCTCGACAGAGGTTGAGAACTCAG 1309  
 114731 TAGACTTACTCCATTGTGCTGTCTCACCAGCACCCCTCGGAAAGGTTGAGAACTCAG 114672  
 1310 AGCCACCCCTTGAGACTGAAGCAAGCGCAAGTGTGTCTGA 1350  
 114671 AGACATCCTGTGAAGCTGACACACACAAAGTATTGTCTGA 114631

AK098608 1609 bp mRNA linear PRI 15-JUL-2002  
 Homo sapiens CDNA FLJ25742 fis, clone TST05965.  
 AK098608  
 AK098608.1 GI:21758666  
 oligo capping: fis (full insert sequence).  
 Homo sapiens testis cDNA to mRNA, clone\_lib:TST clone:TST05965.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,  
 Arita, M., Mueasino, K., Yuuki, H., Hara, H., Suzuki, Y., Hata, H.,  
 Nakagawa, K., Mizuno, S., Morinaga, M., Kawamura, M., Sugiyama, T.,  
 Irie, R., Otsuki, T., Sato, H., Nishikawa, T., Sugiyama, A.,  
 Kawakami, B., Nagai, K., Isogai, T. and Sugano, S.  
 NEDO human cDNA sequencing project.  
 Unpublished  
 2 (bases 1 to 1609)  
 Sugano, S. and Suzuki, Y.  
 Direct Submission  
 Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Laboratory of Genome Structure, Human Genome  
 Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
 (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
 Fax:81-3-5449-5416)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; CDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); CDNA library  
 construction and 5'-end one pass sequencing: Institute of Medical  
 Science, University of Tokyo, Laboratory of Genome Structure, Human  
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for  
 full insert sequencing: RAB and Helix Research Institute.  
 Location/Qualifiers  
 1..1609  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="TST05965"  
 /tissue\_type="testis"  
 /clone\_libs="TST"  
 /note="cloning vector: pME18SFL3"  
 93..1472  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAC05352.1"  
 /db\_xref="GI:21758667"  
 /translation="METVFENEDESTGVSSEKEIVLQRLHLSPFPISTVLYL  
 CEVAFGLYMEIYKANDTFWMSFTIIVGALDQIILMFNKLRRNKAALFEW  
 HTLLGSPVRLCHTIRNYHKMLKNLQKEETQVSTIKRNTMLREIAFSTRDNPMQO  
 KAFKYSVIOALPGSVPOLIQMYLSKIREWPNRALLMTFSLISVTYTGARCNILA  
 TQISNDTDTKLPLIEFFCVMMRLEVISRVVTALFIASLKULSPVLVLIYFVSI  
 LAPLEWFKSGAHLPGKKNENSMVGTVMLEFLITLLYAAINFCSKSAVKQLSDDDKI  
 IDGRGWRGHRIUHYSPQLENVIMILVFRFGKTLNCCDSLIAVQLIISYLLATGFI  
 MLLFYQLYPMQSGKVLPORTENPEAPYVYVNIETKKNKNKQLRNYCNSNRVGYH  
 SIRKSMTCSS"

486 a 290 c 336 g 497 t

BASE COUNT  
 RIGIN

Query Match 32.7%; Score 441.2; DB 9; Length 1609;  
 Best Local Similarity 60.7%; Pred. No. 1.8e-116;

QY 1081 GTGATCATGCTGCTGGTGTGTTTAAAGTCTCTTGGAGTGAAGTGTACTGAATTAATCTGTCAT 1140  
 |||||  
 Db 1158 GTGATATGATATGTTATGTTCTTGGAGGAAACTTTGCTGAATGTTGAC 1217  
 |||||  
 QY 1141 TCCTGTGATGCTGTCAGCTCATTAATGCTTATCTGATTTCCATTTGATCTCATCTCCTT 1200  
 |||||  
 Db 1218 TCATTAATGCTGTCAGCTCATATAAGCTACCTATTGGCCACTGGCTTTATGCTCTC 1277  
 |||||  
 QY 1201 TTCTTCAGTACTTGCATCCATTTGGCTCA 1230  
 |||||  
 Db 1278 TTCTATCATGATTTGTACCCATGGCAGTCA 1307  
 |||||

RESULT 5  
 AF155511 2911 bp mRNA linear ROD 02-DEC-1999  
 LOCUS Mus musculus KX antigen (Xk) mRNA, complete cds.  
 DEFINITION Mus musculus KX antigen (Xk) mRNA, complete cds.  
 ACCESSION AF155511 AF064772  
 VERSION AF155511.1 GI:6502962  
 KEYWORDS  
 SOURCE Mus musculus.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 2911)  
 Collec.E., Collin.Y., Carbonnet.F., Hattab.C., Bertrand.O.,  
 Cartton.J.P. and Kim.C.L.  
 AUTHORS

TITLE Structure and expression of the mouse homologue of the Xk gene  
 JOURNAL Immunogenetics 50 (1-2), 16-21 (1999)  
 MEDLINE 20009522  
 PUBMED 10541802

REFERENCE 2 (bases 1 to 2911)  
 Le Van Kim.C., Collec.E. and Collin.Y.  
 AUTHORS Direct Submission  
 TITLE Submitted (01-JUN-1999) INSERM U76, INTS, Alexandre Cabanel, Paris  
 JOURNAL 75015, France

FEATURES  
 Location/Qualifiers  
 1..2911  
 /organism="Mus musculus"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /map="between Gata1 and Cybb"  
 /sex="male"  
 /tissue\_type="skeletal muscle"  
 /dev\_stage="8-12 weeks"  
 1..2911  
 /gene="Xk"  
 1..141  
 /gene="Xk"  
 142..1482  
 /gene="Xk"  
 /notes="Homo sapiens Kx blood group protein homolog"  
 /codon\_start=1  
 /product="KX antigen"  
 /protein\_id="AA14527.1"  
 /db\_xref="GI:6502963"  
 /translation="MKPPASVIASVFLVFAETAALYLSSTYSAGDRMQVLTLLFS  
 LMPCLVQFTLLFVHRLSDRPLALIMHLQGLYRCCEVFCIVCOSDNEEYVS  
 ITKRQPKDGLSEVEKEVQAEGLITHRSAPSRASVIOAFILGAPOLTLQLYITV  
 LEONITGRCFIMTSLLSIVYGNALRNILAIKIDYEVKVKPLAYVUCIFLWRSFE  
 IATRVVLVFTSVLKIVVAVLVNFFSFLPIWVFCGSGFFPENIEKLSRVGT  
 TVLCPULLYAGINMFCWSAQLKIDNPILISKQNWYRLLIYYMTRFSENSVLLLL  
 WYFKTDIYVVCAPLLILQLLIGCTGLFVYQFPHFKPLFSVSESFRALL  
 RCACWSLSLRKSEPVGRIDTLKACTEQDVMPTTSKVIPEATDIWAVDLCSA"  
 1483..>2911  
 /gene="Xk"

BASE COUNT 741 a 700 c 609 g 861 t  
 ORIGIN

Query Match 22.4%; Score 301.8; DB 10; Length 2911;  
 Best Local Similarity 55.5%; Pred. No. 5.7e-76;

Matches 629; Conservative 0; Mismatches 492; Indels 12; Gaps 2;  
 QY 118 TCACACCTTTTGTACTGTGGGAGCGTGCATCTGCTTTGTACATGTTAGAAATCTATCGA 177  
 |||||  
 Db 169 TCCGTGTCTCTGTTCTGTCGGCCGAGACGCGCGCGCTCTACTTTGAGCAGCACCTACCGC 228  
 |||||  
 QY 178 AAGAATAGTGAACCTTACCAGATGACATACACCTTTCTCTTCTTTATGTTTTCATCCATT 237  
 |||||  
 Db 229 TCCGACGGGACCCGATGTGACAGTGTGACGCTCTCTCTCTCTGATGCCCTCGGCC 288  
 |||||  
 QY 238 ATGCTCCAGTTGACCTCATTTTGTCCACAGAGATCTAGCCAAAGATAAACCGCTATCA 297  
 |||||  
 Db 289 CTGGTGCAGTTTCAGCTCTCTCTCCACCGGACCTCAGCGCGATCGCCCACTGGCG 348  
 |||||  
 QY 298 TTATTATGATCTAATCCCTCTTGGACCTGTTATCAGATGTTTGGAGGCCATGATTAG 357  
 |||||  
 Db 349 CTGCTCATGACCTGCTCCAGCTCGGCCCTGTACAGGTGTGTGAAGTCTTTTGTATC 408  
 |||||  
 QY 358 TACCTCACACTGTGGAAGAAAGAGGAGGAGGAGGAGCCCTATGTGAGCTCACCCGAAA- 416  
 |||||  
 Db 409 TAC-----TGTCACTCCGACCAAGATGAAGAACCTTATGTGAGCATCACTAAGAAA 459  
 |||||  
 QY 417 --GAAGATGCTAATAGATGGCAGGAGTGTCTGATAGAAATGGAGGTGGGCCACTCCATC 474  
 |||||  
 Db 460 CGACAGATGCCAAAGATGGCTTTTCAGAGGAGGTGGAGAAAGAGTTGGCCAGGCAGAA 519  
 |||||  
 QY 475 CGGACCTTGGTATGACCCGCAATGCCCTAAGAGTATGTACAGATCCAGGCTTCCCTG 534  
 |||||  
 Db 520 GGCAAGCTGATTAACCCACCGTTCTGCAATTCAGCCGGGATCAGTGATACAGGCTTTCCTA 579  
 |||||  
 QY 535 GGCTCAGTGGCCCGAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCC 594  
 |||||  
 Db 580 GGCTTGTCCCCACAACTGACCTTAACTGATATATACTGTCTTGGAGCAGAACTACT 639  
 |||||  
 QY 595 CTGGGTAGAGTTGTGCTAATGTTATTTTCCCTGGTATCTGTACCTATGGGGCCACCTT 654  
 |||||  
 Db 640 ACTGGAAGATGCTTTCATCATGACCTTGTCCCTGCTCTATTGTCTATGGCGCTTACGT 699  
 |||||  
 QY 655 TGAATATGTTGGTATTCAGATCAAGTACGATGATCAAGATTCGGCTTGGGGCACA 714  
 |||||  
 Db 700 TGAACATCTTAGCCATCAAAATCAAGTATGATGATGAGGTCAAGTCAAAAGTCAAAACCCCTG 759  
 |||||  
 QY 715 GAAGTCTCTGCAATCACCATCTGGCGGACATTTGAGATCACTTCCCGCTCTCTGATTTCTG 774  
 |||||  
 Db 760 GCTTATGCTGTATCTCTCTCGGAGAGCTTTGAGATTCGACCCGAGTCACTGCTCTG 819  
 |||||  
 QY 775 GTGCTCTTCTCAGCCACTTTGAAATTTGAAGGCTGTGCCCTTCCCTAGTCTCAACTTCTCTG 834  
 |||||  
 Db 820 GTCTCTTTTACCTCTGCTCTGAAAGATCTGGGTGGTGGCAGTCACTACTGTCACACTTCTTC 879  
 |||||  
 QY 835 ATCATCTCTTTGAGCCCTGGATTAAGTTCTGGAGAAAGTGTGGCCAGATGCCCAATTAAC 894  
 |||||  
 Db 880 AGCTTCTTCTTATATCTCTGATCGTTTCTGTTGTCAGTGGCTCCCATTTCCCTGAGAAC 939  
 |||||  
 QY 895 ATTGAGAAAAAATTTACGCCCGGTGCGCACTCTGTTGGTGGTCTGATTTTCAGTCAACATCTC 954  
 |||||  
 Db 940 ATAGAGAGGCCCTTAAGTAGGGTGGGTACCACTATCGTGTCTGCTCTCTCACTTTTACTC 999  
 |||||  
 QY 955 TATCTCGCATCAACTTCTCTTGTGCTGTCAGCTTTGAGTTGAGGTTGGCAGACAGAGAT 1014  
 |||||  
 Db 1000 TATGCGGCATCAACATGTTCTGCTGCTGTCAGTCTGTTTCAAGTGAATAATCGACATCCGGAG 1059  
 |||||  
 QY 1015 CTGCTGACAAAAGGGCAGAACTGGGGACATATGGGCTGCACTATAGTGTGAGGTTGGTA 1074  
 |||||  
 Db 1060 CTATCAGCAAGTCCCAAGAACTGGTACCGCTGCTCATTTTACTATACAGCAGATTCATC 1119  
 |||||  
 QY 1075 GAGATGATGATGCTGTTGTTTAAAGTCTTCTTGGAGTGAAGTGTACTGAATTAC 1134  
 |||||  
 Db 1120 GAGAACTCCGTCCTCTGCTGCTGTTATTTTAAACTGACATATACATGATGTG 1179  
 |||||  
 QY 1135 TGTCACTCTGATGCTTGCAGTCACTATTATGCTTATCTGATTTCCATGACTTCAATG 1194  
 |||||  
 Db 1180 TGTGCGGCTCTGTTGATCTTGCAGCTGCTTATTTGGTACTGCACTGGCACTTCTTCTCATG 1239  
 |||||

y 1195 CTCCTTTCTCCAGTACTGTCATCTGCGCTCAGCTTCCACCATTAATGT 1247  
 |||||  
 b 1240 CTTGTGTTCTATCAGTTTTTCCACCCCTTGCAAAAAGCTCTTCTCTCCAGTGT 1292  
 |||||  
 RESULT 6  
 X332514  
 OCUS AX332514 5096 bp DNA linear PAT 09-JAN-2002  
 DEFINITION Sequence 3023 from Patent WO0194629.  
 CCSSION AX332514  
 ERSION AX332514.1 GI:18123148  
 EYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
 Horrigan, S., Soppet, D.R. and Weaver, Z.  
 TITLE Cancer gene determination and therapeutic screening using signature  
 gene sets  
 JOURNAL Patent: WO 0194629-A 3023 13-DEC-2001;  
 Avalon Pharmaceuticals (US)  
 FEATURES  
 Location/Qualifiers  
 source  
 1. 5096  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 1392 a 1064 c 1022 g 1618 t  
 RIGIN  
 Query Match 21.5%; Score 289.6; DB 6; Length 5096;  
 Best Local Similarity 54.6%; Pred. No. 2.2e-72;  
 Matches 627; Conservative 0; Mismatches 509; Indels 12; Gaps 2;  
 y 118 TCACCTTTTCTACTGTGGGAGGCTGCATCTCTCTTGTACATGTTAGAACTATCA 177  
 |||||  
 b 110 TCCGTGTTCTGTTGCGCCGAGACAACGGCGGCGCTCAGCTGAGCAGCACTACCGC 169  
 |||||  
 y 178 AAGATAGTGAACATTACCGGATGACATACACCTTTTCTTTTATGTTTTCATCCATT 237  
 |||||  
 b 170 TCGGGGGGGGACCGCATGTGGCAGGCGCTGACGTGCTTTTCTGCTACTGCTTGGCGG 229  
 |||||  
 y 238 ATGGTCCAGTTCACCTCATTTTGTCCACAGAGTCTAGCCAAAGATAAACCGCTATCA 297  
 |||||  
 b 230 CTCGTGCACTCAGCTTCTCTCTGACACCGGACCTCAGCGCGACCGCGCTCGTA 289  
 |||||  
 y 298 TTATTTATGCATCTAATCCTCTTGGACCTGTTTATCAGATGTTTGGAGGCCATGATTAG 357  
 |||||  
 b 290 CTGCTGCTGCACCTGCTGCAACTTGGGCCCCCTTTTCAGGTGTTTGAAGTCTTCTGCATC 349  
 |||||  
 y 358 TACCTCACACTGTGGAGAAAGAGGAGGAGGAGCCCTATGTCAGCTCACCAGAAAG 417  
 |||||  
 b 350 TACTTTC-----AGTCAGGCAACAATGAAGAGCCTTATGTCAGTATCACCAAGAG 400  
 |||||  
 y 418 AAG---ATGCTAATAGTCGAGGAGGTGCTGATAGATGGGAGGTGGCCACTCCATC 474  
 |||||  
 b 401 AGGCAATGCCAAAATAATGGCTCTCAGAGGAGATTGAGAAGAGGTGGGCCAGGAGAA 460  
 |||||  
 y 475 CGGACCTGGCTATGCACCGCAATGCTTACAAACGTATGTACAGATCCAAAGCCTTCCTG 534  
 |||||  
 b 461 GGCAACTAATCACCCACCGATCAGGTTTACGGCGGCTCGGTGATCCAGGCTTCTTG 520  
 |||||  
 y 535 GGTCTAGTCCCGCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGAGAGGTTCCT 594  
 |||||  
 b 521 GGCTCAGCCCCCAGCTGACCCCTACAGCTGTACATAAGTGTATGTCAGCAGGACGTCAC 580  
 |||||  
 y 595 CTGGGTAGAGTTGTGTATGTTTCCCTGGTCTGTCTACCTATGGGCGCACCTT 654  
 |||||  
 b 581 GTTGGGAAGAGTCTCTCATGACCATCTCTGTGTCATTTGTGATGGAGCTTGGCG 640  
 |||||  
 y 655 TGAATATGTTGGCTATCCAGATCAAGTACGATGACTACAAAGATTGCGCTTGGGCCACTA 714  
 |||||

Db 641 TGCACATCTCTAGCCATCAAAATCAAGTACGATGAGTATGAAGTCAAAAGTGAAGCCTCTG 700  
 Qy 715 GAAGTCTCTGATCACCATCTGGGACATTTGAGATCACTTCCGCTCTCTGATTTCTG 774  
 Db 701 GCCTATGTCTGTATCTTCTCTGGAGGAGCTTTGAGATTCGACATCGAGTTGTAGTCTG 760  
 Qy 775 GTGCTCTTCTCAGCCACATTTGAAATTTGAAGCTGTGCGCTTCTCTAGTCTCAACTTCTG 834  
 Db 761 GTCTCTTTTACTCTCGTCTGAGACCTGGGTGGTGTATTAATACTCATCACTTCTTC 820  
 Qy 835 ATCAPCTCTTTGAGCCCTGGAATTAAGTTCTGGAGAAATGTGGCCAGATGCCCAATTAAC 894  
 Db 821 AGTTTCTTCTTGTACCTCCCTGGATCTCTTCTGTGTGACGTGGTTCGCCATTCCTCGAGA 880  
 Qy 895 ATTGAGAAAACCTTCAGCGGGTGGGCACTCTGGTGTCTCTGATTTTCACTCACCATCTCT 954  
 Db 881 ATAGAGAAGGCCCTCAGTAGAGTGGGCAACCACTTGTACTATGCTTTCTAACTTTTACTC 940  
 Qy 955 TATGTGGCATCAACTTCTCTTCTGTGTGCTCAGTTTGTGAGTTGAGGTGGCAGACAGAGAT 1014  
 Db 941 TATACTGTATCAACATGTTCTGTGTGTCTGTGTGCTGCTGACAGCTGAAAATTTGACAGCCCTGAC 1000  
 Qy 1015 CTCGTGCAAAAGGCGCAACTGTGGGACATATGGGCCCTGCACTATAGTGTGAGGTGGTA 1074  
 Db 1001 CTCATCAGCAAGTCCCAATAATTGTGTAACCACTACTGTGTGTATTACATGATAAGATTATC 1060  
 Qy 1075 GGAATGTGATCATGTTCTTGGTTTAAAGTTCTTGGTGAAGAGTGTACTGAATTAC 1134  
 Db 1061 GAGATGCCATCTCTCTCTCTCTGTGTGTATCTTTTCAAGACTGACATCTATATGATGTG 1120  
 Qy 1135 TGTCTATCTTCTGATTGCTTGCAGCTCATTTATGTTCTTATCTGATTTCCATGACTTTCATG 1194  
 Db 1121 TCGGCACTCTGTGTGTCTGCACTGCTCATTTGGGTACTGCAAGCCATCTCTTCTCATG 1180  
 Qy 1195 CTCCTTTTCTTCCAGTACTTGCATCCNTTGGCTCCTCTTCCACCCATAATGATAGTAGAC 1254  
 Db 1181 CTGTATTCTATCAGTCTTCTTCCACCTTTCGAAAAAGCTTTTCTTCCAGTGTCTTCTGAA 1240  
 Qy 1255 TACCTCCA 1262  
 Db 1241 GGCTTTCA 1248  
 RESULT 7  
 HSXKMP  
 LOCUS Homo sapiens mRNA for membrane transport protein (XK gene). PRI 19-MAY-1999  
 DEFINITION  
 ACCESSION Z32684  
 VERSION Z32684.2 GI:4883432  
 KEYWORDS membrane transport protein; XK gene.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 5096)  
 AUTHORS Ho, M., Chelly, J., Carter, N., Danek, A., Crocker, P. and Monaco, A.P.  
 TITLE Isolation of the gene for McLeod syndrome that encodes a novel  
 membrane transport protein  
 JOURNAL Cell 77 (6), 869-880 (1994)  
 MEDLINE 94273191  
 PUBMED 8004674  
 REFERENCE 2 (bases 1 to 5096)  
 AUTHORS Ho, M.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-APR-1994) Meng F Ho, Human Genetics, Imperial Cancer  
 Research Fund, Institute of Molecular Medicine John Radcliffe  
 Hospital, Headington, Oxford, OX3 9DU, United Kingdom  
 REMARK revised by [3]  
 REFERENCE 3 (bases 1 to 5096)  
 AUTHORS Ho, M.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-OCT-1997) Meng F Ho, Human Genetics, Imperial Cancer  
 Research Fund, Institute of Molecular Medicine John Radcliffe

Query Match	21.5%	Score	289.6	DB	9	Length	5096
Best Local Similarity	54.6%	Pred. No.	2.2e-72				
Matches	627	Conservative	0	Mismatches	509	Indels	12
Gaps							
Qy	118	TCACACCTTTTGTACTGTGGGAGGCTGCATCTCGTTTGTACATGGTTAGAAATCTATCGA	177				
Db	110	TCGGTGTTCCTGTTCGTGCGCGAGACAACGGCGGGCTCAGCCTGAGCAGCACCTACGCG	169				
Qy	178	AAGAATAGTGAACCTTACCGGATGACATACACCTTTTCTTTTCTTTATGTTTTCATCCATT	237				
Db	170	TCGGCGGGGACCGCATGTGGCAGCGGTGACGTTGCTTTTCTCGCTACTGCTTGGCGG	229				
Qy	238	ATGGTCAGTTGACCCCTCATTTTGTGCACAGAGATCTAGCCAAAGATAAACCCCTATCA	297				
Db	230	CTCGTGCAGCTCAGCTTCTCTTCGTACACGCGACCTCAGCCGCGACCGCCGCTCGTA	289				
Qy	298	TTATTTATGCATCTAAATCCTCTTGGGACCTGTTATCAGATGTTTGGAGGCGATGATTAAAG	357				
Db	290	CTGCTGTGTGCACTGCTGCAACTTGGGCCCCCTTTTCAGGTGTTTTGAAGTCTTCTGCATC	349				
Qy	358	TACCTCACACTGTGGAGAGAGAGGACAGAGAGAGCCCTATGTACGCTCACCCGGAAG	417				
Db	350	TACTTTC-----AGTCAGGCAACATGAAGAGCCTTAATGTCAGTATCACCAGAAG	400				
Qy	418	AAG---ATGCTAAATAGATGGCGAGAGGTGCTGATAGAAATGGGAGGTGGGCCACTCCCATC	474				
Db	401	AGGCAATGCCCCAAAATGGCCCTCTCAGAGGAGATTGAGAGGAGGTGGGCCACGCGAGAA	460				
Qy	475	CGGACCTGGGTATGCACCCGAATGCCTCAAAACGTATGTACAGATCCAAAGCCTTCGTG	534				
Db	461	GGCAAACTTAATCACCCACCCGATCAGCGTTTACGCCGGCGCTCGGTGATCTCAGGCTTCTTG	520				

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
304..1536
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD34798.1"
/db_xref="GI:21438822"
/translation="MQQALTLFSLPCALVQLTLLFVHRDLSRDLRVLLHLHLQLG
PLFCFEVFCYFQSGNNEEYYSIKRQMPQKSLSEEKEVGGQAGKLIHRSAF
SRASVIQAFGLAQPLQLQIYSVMOQDVGRLSLMTISLLSVIAGRCNLIAIKI
KYDEWGVKPLAVLCVFLNRSFIATRVVVLVLTFSVLTWVVIILINFFSFFLYK
WILFCSQSPPENIEKALSRVGTIVLCPLTLYLTGINMFCVAVQLKIDSPFLISK
SHNYSVQLLVYMYRIENAILLALWYLEKTDIYMYVCAPIVLQLIAGVCTALLFMV
FYQFHPCKZLGFSSSSGEGFQWRUFCFNCACRQKQCEPIGKEDLQSSRDRDETPSSS
KTSPEQGQFLNADLCSA"
1458 a 1058 c 1035 g 1664 t

```

Query Match	20.9%	Score 282.8	DB 6	Length 5215
Best Local Similarity	55.7%	Prod. No. 2e-70		
Matches 589	Conservative	0	Mismatches 457	Indels 12
Gaps				
208	ACCTTTCTCTCTTTATGTTTTCATCCCAATTATGGTCAGTTTGACCTCATTTTGTGCCAC	267		
319	AGTTGCTTTTCTCGCTACTGCTTGGCGCTCGTGCAGCTCACGCTTCTTCTGTACAC	378		
268	AGAGATCTAGCCAAAGATAAACCGCTATCATTTATTTATGCACTTAATCCTCTTGGGACCT	327		
379	CGGACCTTCAGCGGACCGCCGCTCGTACTGCTGCTGCACTCTGCTCAACTTGGGCC	438		
328	CTTATCAGATGTTTGGAGGCCATGATTAAGTACTCTACACTGTGGAAGAAAGAGGAGCAG	387		
439	CTTTTCAGGTGTTTTGAAGCTCTTGCACTACTTTC-----AGTCAGGCAACAAT	489		
388	GAGGAGCCCTATGTCAGCTCACCCGAAAGAG---ATGCTAATAGATGGCGAGGAGTG	444		
490	GAAAGCCTTATGTCAGTATCACCAAGAGAGGCCAAATGCCCAAATAATGGCTCTCACAG	549		
445	CTGATAGAATGGAGGTGGGCCACTCTCCATCCGAGCCCTGGCTATGACCGCAATGCCTAC	504		
550	GAGATTGAGAAGAGGTGGGCCAGGAGGAGCAAACTAATCACCCACCGATCAGCGTTC	609		
505	AAACGATATGTCAGAGATCAAAGCTTCTGGGCTCAGTCCGCCAGCTGACCTATCAGCTC	564		
610	AGCCGGGCTGGTGATCAGAGCTTCTTGGGCTCAGCCCCCAGCTGACCCCTACAGTGTG	669		
565	TATGTGAGCCTGATCTCTGCAGAGGTTCCCCCTGGGTAGAGTTGTGCTAATGGTATTTTCC	624		
670	TACATAAGTGTATGTCAGCAGGAGCGTCACTGTTGGAAGAAGTCTCTCATGACCATAACC	729		
625	CTGGTATCTGTCACTTATGGGGCCACCCTTTGCAATATATGTTGGCTATCCAGATCAAGTAC	684		
730	CTGTTGTCCATTGTATGGAGGCTTGGCCTCGCAATCCTAGCCATCAAAATCAAGTAC	789		
685	GATGACTTACAAGATTCCGCTTGGGCCACTAGAAAGTCTCTGCATCACCATCTGGCGGACA	744		
790	GATGAGTATGAAGTCAAGTGAAGCCTCTGGCCATATGCTGTATCTTCTGTGGAGGAGC	849		
745	TTGGAGATCACTTCCCGCCCTCTGATTTCTGGTGCTCTTCTCAGCCACTTTGAAATTTGAAG	804		
850	TTTGGATTGCCACTCGAGTTGTAGTCTGGTCTCTTTTACCTCCGTCTCGAAGACCTGG	909		
805	GCTGTGCCCTTCTAGTGTCTCACTTCTCTGATCATCTCTTTGAGCCCTGGATTAAGTTC	864		
910	GTGGTGGTTATAATACTCATCAACTTCTTTCAGTTTCTTCTGTGACCCCTGGATCTCTTTC	969		
865	TGAGAAAGTGGTGCCAGATGCCCAATAACATTGAGAAAACCTTCAGCCGGGTCCGCACCT	924		
970	TGTTGAGTGGTTTCCCATTTCCCTGAGACATAGAGAAAGGCCCTCAGTAGAGTGGGCACC	1029		
925	CTGGTGGTCTGATTTTCAAGTACCACTCTCTATATGCTGGGCAATCAACTTCTTTGCTGGTCA	984		
1030	ACCAATTGTAATGCTTCTTAACTTTTACTCTATACTGGTATCAACAATGTTCTGCTGGTCT	1089		

Qy	985	GCCTTGCAGTTGAGTTGGCAGACAGAGATCTCGTCGACAAAGGGCAGAACTGGGACAT	1044
Db	1090	GCTGTACAGCTGAAATTGACAGCCCTGCAGCTCATCAGCAAGTCCCATTAATTGGTACCAG	1149
Qy	1045	ATGGGCTGCACATATAGTGTGAGGTTGGTAGAGAAATGTGATCATGGTCTTGGTCTTTTAAG	1104
Db	1150	CTACTGGTGTATTACATGATAAGATTTCATCGAAGTGCATCTCTCTCTCTGGTAT	1209
Qy	1105	TTCTTTGGAGTGAAAGTGTTCGAAATTAATCTGTCATCTCTTGTGATTCGCTTGCAGCTCAT	1164
Db	1210	CTTTTCAAGACTGACATCTATATGTATGTGTGGCCACTCTGTGTGGTCTCTGCAGCTGCTC	1269
Qy	1165	ATTGCTTATCTGATTTCCATTGACTTCATGCTCTTTTCTTCAGTACTTGCATCCATTG	1224
Db	1270	ATTGGGTACTGCACAGCAATCTCTTCATGCTGTATCTATCAGTTCCTTCCACCTTGC	1329
Qy	1225	CGTCACTCTTCCACCAATAATAGTAGACTACTCTCA	1262
Db	1330	AAAAGCTCTTTCTTCCAGTGTTCTGAAGGCTTTCA	1367
RESULT	9		
AC005301/c			
LOCUS		113688 bp DNA linear	PRI 27-MAY-2000
DEFINITION		Homo sapiens chromosome 22q11 clone pl5j16, complete sequence.	
ACCESSION		AC005301	
VERSION		AC005301.22	GI:7107554
KEYWORDS		HTG.	
SOURCE		Homo sapiens.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		1 (bases 1 to 113688)	
JOURNAL		Cantu, L.A., McDermid, H. and Roe, B.A.	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 113688)	
TITLE		Cantu, L.A., McDermid, H. and Roe, B.A.	
JOURNAL		Direct Submission	
REFERENCE		Submitted (18-JUL-1998) Department of Chemistry And Biochemistry,	
AUTHORS		The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
TITLE		OK 73019, USA	
JOURNAL		3 (bases 1 to 113688)	
REFERENCE		Cantu, L.A., McDermid, H. and Roe, B.A.	
AUTHORS		Direct Submission	
TITLE		Submitted (16-OCT-1999) Department of Chemistry And Biochemistry,	
JOURNAL		The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
REFERENCE		OK 73019, USA	
AUTHORS		4 (bases 1 to 113688)	
TITLE		Cantu, L.A., McDermid, H. and Roe, B.A.	
JOURNAL		Direct Submission	
REFERENCE		Submitted (28-FEB-2000) Department of Chemistry And Biochemistry,	
AUTHORS		The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
TITLE		OK 73019, USA	
JOURNAL		5 (bases 1 to 113688)	
REFERENCE		Cantu, L.A., McDermid, H. and Roe, B.A.	
AUTHORS		Direct Submission	
TITLE		Submitted (25-MAY-2000) Department of Chemistry And Biochemistry,	
JOURNAL		The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
REFERENCE		OK 73019, USA	
AUTHORS		6 (bases 1 to 113688)	
TITLE		Cantu, L.A., McDermid, H. and Roe, B.A.	
JOURNAL		Direct Submission	
REFERENCE		Submitted (27-MAY-2000) Department of Chemistry And Biochemistry,	
AUTHORS		The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
TITLE		OK 73019, USA	
JOURNAL		7 (bases 1 to 113688)	
REFERENCE		Cantu, L.A., McDermid, H. and Roe, B.A.	
AUTHORS		Direct Submission	
TITLE		Submitted (27-MAY-2000) Department of Chemistry And Biochemistry,	
JOURNAL		The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
REFERENCE		OK 73019, USA	
COMMENT		On Feb 28, 2000 this sequence version replaced gi:6056215.	
		Because these overlapping clones came from different libraries	
		there are numerous instances of insertions, deletions, and single	
		nucleotide polymorphisms in the overlapping regions below.	
		(AC005301[p15j16]) 77414 113688 (0) overlaps AC007064[p8708] 1 36275	
		(800548).	

FEATURES		Location/Qualifiers	
source	1..113688	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="22q11"	
		/clone="p15j16"	
BASE COUNT	37158 a 22385 c 21889 g 32256 t		
ORIGIN			
Query Match 20.5%; Score 276.4; DB 9; Length 113688;			
Best Local Similarity 64.9%; Pred. No. 2.4e-68;			
Matches 409; Conservative 0; Mismatches 221; Indels 0; Gaps 0;			
QY	601	AGAGTTGTGCTAATGTTATTTCCCTGGTATCTGTCACTATGGGCCACCCCTTGCAT	660
Db	101529	ATAGCATGCTGATGACATTTTCCCTGTTATCAGTTACTTATGGGCCATTCGCTGCAAT	101470
QY	661	ATGTGTGCTATCCAGATCAAGTAGCTACTACAAGATTGCGCTTGGGCCACTAGAAGTC	720
Db	101469	ATACTGCCATCCAGATCAGCAATGATGATACCTAACCTAACCTACGCCGATAGAATTC	101410
QY	721	CTCTGCATCACCATCTGGCGACATTTGGAGATCACTTCCCGCTCTGTGATTTCTGTGCTC	780
Db	101409	TTCTGTGCTGATGTGGCGCTTTTGGAGGTTATCTCAGCTGTAGTACTCTGGCAATTT	101350
QY	781	TTCTCAGCCACTTTGAATTTGAAGCTGTGCCCTTCTCTAGTGTCACTTCCTGATCATC	840
Db	101349	TTCAATGCATCTCTGAACATGAAGAGCTACCCGTTTGTGTAATCATATATTTGTATCA	101290
QY	841	CTCTTTGAGCCCTGGATTAAGTTCTCGAGAAGTGGTGCCAGATGCCCAATAACATTTAG	900
Db	101289	TTGTTGACCGTGGCTGGAGTTTGGAAAGTGGAGCTCATCTTCTGGCAACAAGAA	101230
QY	901	AAAACTTCAGCCGGTGGGACCTCTGGGTCTGTGATTTCACTCACCCTCTATGCT	960
Db	101229	AATAATTTCCAATATGGTGGGTACAGTACTGATGCTTCTTGTGATCACACTGCTATGCT	101170
QY	961	GGCATCACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1020
Db	101169	GCCATCACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	101110
QY	1021	GACAAAGGGCAGAACTGGGACATATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1080
Db	101109	GACGGGAGACAGAGTGGGCCATAGATCTCTACACTACAGCTTTCAGTTTGTAGAAAT	101050
QY	1081	GTGATCATGTCTTGTGTTTAAAGTTCTTTGGAGTGAAGTGTACTGAATTAAGTGTAT	1140
Db	101049	GTGATAATGATATTTGTTATTTAGTCTTTTGGAGGAAACTTTGCTGTAATTTGTTGAC	100990
QY	1141	TCCTTGATTCCTTGCAGCTCATTTGCTTATCTGATTTCCATTGATTCATGCTCCTT	1200
Db	100989	TCATTAATTCGGTGCAGCTCATCATAGTACTTATTTGGCCACTGGCTTTATGCTCCTC	100930
QY	1201	TTCTTCCAGTACTTGCATCCATTCGCTCA	1230
Db	100929	TTCTATCAGTATTTGTACCACTGCGAGTCA	100900
RESULT 10			
AC007064/c	124823 bp	DNA	linear
LOCUS			PRI 27-MAY-2000
DEFINITION	Homo sapiens Chromosome 22q11 PAC Clone p8708 in CES region, complete sequence.		
ACCESSION			
VERSION	AC007064		
KEYWORDS	HTG.		
SOURCE			
ORGANISM	Homo sapiens.		
	Homo sapiens.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 124823)		
AUTHORS	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
TITLE	Homo sapiens Chromosome 22q11 PAC Clone p8708 in CES Region		
Unpublished			
2 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (11-MAR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
3 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (26-OCT-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
4 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (27-OCT-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
5 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (05-NOV-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
6 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (06-NOV-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
7 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (29-FEB-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
8 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (25-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
REFERENCE			
9 (bases 1 to 124823)			
REFERENCE	Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.		
AUTHORS	Direct Submission		
TITLE	Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
JOURNAL			
COMMENT			
On Feb 29, 2000 this sequence version replaced gi:6249691. Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below. AC005301(p15j16) 77414 113688 (0) overlaps AC007064(p8708) 1 36275 (88548) AC007064(p8708) 52616 124823 (0) overlaps AC006548(p20K14) 1 62199 (137950).			
1 62199 (137950).			
Location/Qualifiers			
1..124823			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="22q11"			
/clone="p8708"			
41074 a 23932 c 23324 g 36493 t			
BASE COUNT			
ORIGIN			
Query Match 20.5%; Score 276.4; DB 9; Length 124823;			
Best Local Similarity 64.9%; Pred. No. 2.4e-88;			
Matches 409; Conservative 0; Mismatches 221; Indels 0; Gaps 0;			
QY	601	AGAGTTGTGCTAATGTTATTTCCCTGGTATCTGTCACTATGGGCCACCCCTTGCAT	660
Db	24116	ATAGCATGCTGATGACATTTTCCCTGTTATCAGTTACTTATGGGCCATTCGCTGCAAT	24057

Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,  
Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,  
Lorado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maneshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Mosser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokkenwo, S., Oguh, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Picken, R., Primus, B., Pul, L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczky, R., Woodden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 168918)  
Worley, K.C.

Direct Submission  
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 168918)  
Worley, K.C.

Direct Submission  
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:19718450.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GTLM  
Center clone name: CH230-139K19  
----- Summary Statistics  
Sequencing vector: Plasmid,  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap; version 0.950329  
Consensus quality: 104166 bases at least Q40  
Consensus quality: 113489 bases at least Q30  
Consensus quality: 119214 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 74 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1568: contig of 1568 bp in length  
\* 1569 1688: gap of unknown length  
\* 1699 2687: contig of 1019 bp in length  
\* 2688 2787: gap of unknown length  
\* 2788 4137: contig of 1350 bp in length  
\* 4138 4237: gap of unknown length  
\* 4238 5268: contig of 1031 bp in length  
\* 5269 5369: gap of unknown length  
\* 5369 6665: contig of 1297 bp in length  
\* 6665 6765: gap of unknown length



## JOURNAL

Submitted (09-MAY-2001) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

3 (bases 1 to 160200)

REFERENCE  
AUTHORS

Montgomery, K.T., Grills, G., Li, L., Chiu, D., Decker, J., Fusina, M., Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

TITLE  
JOURNAL

Direct Submission  
Submitted (06-JUL-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA

REFERENCE  
AUTHORS

4 (bases 1 to 160200)  
Montgomery, K.T., Grills, G., Li, L., Chiu, D., Decker, J., Fusina, M., Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

TITLE  
JOURNAL

Direct Submission  
Submitted (02-AUG-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA

## MMENT

On Aug 2, 2002 this sequence version replaced gi:21700544.

-----Genome Center:

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcg.org/Sequence/mouse.html>

Contact: hpgc@emdel.mgh.harvard.edu

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550).

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST and cDNA sequences in Unigene. Genes demonstrate at least two exons flanked by consensus splice sites that maintain sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double stranded sequence for all regions. All sequence is completed to a standard of coverage with a minimum of 3 reads with no ambiguities. If the sequence coverage for a region does not meet this standard, it is indicated in the annotation as Low Coverage. Low coverage linkages are verified by PCR product size verification or verification of forward and reverse reads from clones which span the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated average error rate is less than 1 per 10,000 bases using the Consed quality parameters. Regions that do not meet this requirement are annotated as Low Quality.

## -----Summary Statistics

Center project name: AGA

Sequencing vector: pUC18, L08752, 99%; pSMART, AF399742, 1%

Chemistry: Dye-terminator Big Dye, 100%

Assembly program: Phrap version 0.990319

Contig length: 160200

Fraction of Phrap value < 40: 0.000736

Error Rate in Consed: 0.00 per 10,000 bases

Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

```

1000
 900
 800
 700
  #
    bases 600

```

FEATURES									
Source									
-----									
Location/Qualifiers									
1. 160200									
/organism="Mus musculus"									
/strain="C57BL6/J"									
/db_xref="taxon:10090"									
/chromosome="X"									
/clone="RP23-64A09"									
64..91									
repeat_region	/rpt_family="AT-rich"								
346..401									
repeat_region	/rpt_family="B1P"								
complement(739..841)									
repeat_region	/rpt_family="PB1D10"								
790..14058									
gene	/gene="U43384 (Mm.200362) Mus musculus gp91phox (Cybb)"								
join(790..1592,2541..2633,5328..5581,7564..7726,									
9595..9741,12255..12379,13931..14058)									
mRNA	/gene="U43384 (Mm.200362) Mus musculus gp91phox (Cybb)"								
/product="U43384 (Mm.200362) Mus musculus gp91phox (Cybb)"									
1331..1338									
repeat_region	/rpt_family="ID B1"								
complement(1670..1784)									
repeat_region	/rpt_family="MTD"								
complement(3847..3917)									
repeat_region	/rpt_family="tRNA-Ala-GCY"								
complement(4104..4360)									
repeat_region	/rpt_family="L1MA6"								
4825..4971									
repeat_region	/rpt_family="B1P"								
6025..6112									
repeat_region	/rpt_family="PB1D10"								
6119..6169									
repeat_region	/rpt_family="ID4"								
complement(6193..6556)									
repeat_region	/rpt_family="RMER4A"								
6841..6818									
repeat_region	/rpt_family="(GAAA)n"								
6853..6895									
repeat_region	/rpt_family="(CAAG)n"								
6896..7047									
repeat_region	/rpt_family="RSINE1"								
7096..7290									
repeat_region	/rpt_family="B3"								
8042..8092									
repeat_region	/rpt_family="(CA)n"								
complement(8098..8167)									
repeat_region	/rpt_family="ID4"								
complement(8372..8503)									
repeat_region	/rpt_family="B1P"								
complement(9054..9070)									
repeat_region	/rpt_family="B3A"								
9071..9104									
repeat_region	/rpt_family="(TC)n"								
complement(9105..9282)									
repeat_region	/rpt_family="B3A"								
9366..9467									
repeat_region	/rpt_family="PB1D7"								
complement(10128..10335)									
repeat_region	/rpt_family="MER2B"								
complement(10406..10609)									
repeat_region	/rpt_family="MIR"								
10830..11215									
repeat_region	/rpt_family="MTE"								

QY	1106	TCTTTGGAGTAAAGTGTACTGAATTAATCTCTGATTCCTGATTCCTGAGCTCATTA	1165
Db	142858	TTTTTAAACTGACATATACATGATGTGTGCGGCTCTGTGATCTTGGCAGCTGCTTA	142799
QY	1166	TTGCTTATCTGATTTCCCATGATTCATGCTCTCTTTTCTTCCAGTACTTGCATTCATGC	1225
Db	142798	TTGGGTACTGCACTGCGCATCTCTTCATGCTGTGTCTTATCAGTATTTTCCACCTTGCA	142739
QY	1226	GCTCACTCTTCCACCATATATG	1247
Db	142738	AAAGCTCTTCTCTCCAGTGT	142717
RESULT	13		
LOCUS	AL672060		
DEFINITION	Mus musculus chromosome X clone RP23-423L11, *** SEQUENCING IN		
PROGRESS	***, in ordered pieces.		
ACCESSION	AL672060		
VERSION	AL672060.8	GI:21732098	
KEYWORDS	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.		
TITLE	Tracey, A.		
JOURNAL	Direct Submission		
COMMENT	Submitted (09-JUL-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgehire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 11, 2002 this sequence version replaced gi:21531232.		
Center:	Wellcome Trust Sanger Institute		
Center code:	SC		
Web site:	http://www.sanger.ac.uk		
Contact:	humquery@sanger.ac.uk		
-----	Project Information		
Center project name:	BM423L11		
-----	Summary Statistics		
Assembly program:	XGAP4; version 4.5		
Sequencing vector:	plasmid; L08752; 100% of reads		
Chemistry:	Dye-terminator; 5% of reads		
Chemistry:	Dye-terminator Big Dye; 94% of reads		
Consensus quality:	200633 bases at least Q40		
Consensus quality:	200855 bases at least Q30		
Consensus quality:	200682 bases at least Q20		
Insert size:	200698; sum-of-contigs		
Insert size:	197898; 5.1% error; agarose-fp		
Quality coverage:	10.93x in Q20 bases; sum-of-contigs Quality coverage: 11.26x in Q20 bases; agarose-fp		
-----			
* NOTE:	This is a 'working draft' sequence.		
* This	sequence will be replaced		
* by	the finished sequence as soon as it is available and		
* the	accession number will be preserved.		
FEATURES	Location/Qualifiers		
source	1..200698		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/chromosome="X"		
	/clones="RP23-423L11"		
	/clone_lib="RPC1-23"		
misc_feature	1..200698		
	/note="assembly_fragment:03870"		
	clone_end:T7		
	vector_side:right"		
BASE COUNT	58266 a 41853 c 40587 g 5992 t		
ORIGIN			
Query Match	16.1%; Score 218; DB 10; Length 160200;		
Best Local Similarity	57.5%; Pred. No. 2.2e-51;		
Matches	392; Conservative 0; Mismatches 290; Indels 0; Gaps 0;		
QY	566	ATGTGACCTGATCTCTGACAGAGTTCCTCGGTGAGTGTGCTAATGTTATTTCCC	625
Db	143398	AGGTGAGATTGACAGTGTCTTCTCTCTCTTCAGGCTTCATCATGACCTGTGCC	143339
QY	626	TGTTATCTGACCTATGTTGGGCGACCTTTTGCATATGTTGGCTATCCAGATCAAGTACG	685
Db	143338	TGCTGTCTATGTTATGTTATGTTGGGCGCTTACGTTGCAACATCTTAGCCATCAAAATCAAGTATG	143279
QY	686	ATGACTACAAGATTCGCTTGGGCGCACTAGAGTCTCTGCAATCCATCTGCGGACAT	745
Db	143278	ATGATGATGAGTCAAAAGTCAACCCCTGGCTTATGCTGTATCTCTCTGAGAGGCT	143219
QY	746	TGGAGATCACTTCCCGCTCTGATTTGGTGTCTTCTCAGGCACCTTTGAAATGAAGG	805
Db	143218	TTGAGATTGACCCGAGTCACTGCTGCTCTTTTACCTCTGCTGCAAGATCTGGG	143159
QY	806	CTGTGCTTCTGATGCTCACTTCTGATCATCTCTTTAGCCCTGATTAAGTTCT	865
Db	143158	TGGTGGCAGTCATCTCGTCAACTTCTTCACTTCTTATATCTGATCTGATCTGTTCT	143099
QY	866	GGAGAAGTGTGCGGACATGCCAATAACATGAGAAAACTTTCAGCGGCTCGGCATC	925
Db	143098	GGTGCAGTGGCTCCCATCTCTGAGAACATAGAGAGGCCCTTAAGTAGGGTGGGTACCA	143039
QY	926	TGTTGGTCTGATTTTCACTCACCATCTCTATGCTGGCATCAACTCTCTTCTGCTGCTCAG	985
Db	143038	CCATCGTCTGCTGCTTCTCTCACTTCTATCTATCGGCAATCAATGTTCTGCTGCTCAG	142979
QY	986	CTTTGCAAGTGTGGTGGCAGACAGATCTCTGCAAGGCGGAGAACTGGGGACATA	1045
Db	142978	CTGTTCACTGCTGAAATCGCAATCCGAGGCTCATCGCAAGTCCCGAAGTGGTACCGCC	142919
QY	1046	TGGGCTTGCATATGTTGTCAGGTTGTGACAGATGATCATGGTCTTGGTTTTAAGT	1105
Db	142918	TGCTCATTTACTATCAGCAGATTCATCGAGAACTCCGCTCTCTGCTCTCTGCTGATTT	142859

Matches	392;	Conservative.	0;	Mismatches	290;	Indels	0;	Gaps	0;
/	566	ATGTGAGCCTGATCTCTGCAGAGAGTTCCCGTGGTAGACTGTGCTAAATGGATTATTC	625						
/	119410	AGGTGAGATTGACAGTGCTCTTCTCTCTCTCTCTCTCTCTCATCAGCCTTGTTGCC	119469						
/	626	TGGTATCTGTCAACCTATGGGGCCACCCTTTGGAATATGTTGGCTATCCAGATCAAAGTACG	685						
/	119470	TGCTGTCTATTGTCTATGGCGCTTACGTTGCAACATCTTAGCCATCAAAATCAAAGTATG	119529						
/	686	ATGACTACAAGAATTGCCCTTGGCCCACTAAGAAGTCTCTGCAATCACCACTGCGCGGACAT	745						
/	119530	ATGAGTATGAGGTCAAAGTCAAAACCCCTGGCTATGTCTGTATCTTCTCTGGAGAGCT	119589						
/	746	TGGAGATCACTTCCCGCCTCTCTGATTTCTGGTGCTCTTCTCAGCCACTTTGAAATTTGAAGG	805						
/	119590	TTGAGATTGCCACCGGAGTCATCGTCTGGTCTCTTTACCTCTGCTCCTGAAGATCTGGG	119649						
/	806	CTGTGCCCTTCTTAGTGCTCAACTTCTCTGATCATCTCTTTGAGCCCTGGATTAAGTTCT	865						
/	119650	TGTTGGCAGTCACTACTCGTCAACTTCTTTCAGCTTCTTCTTATATCCCTGGATCGTTTTCT	119709						
/	866	GGAGAAGTGGTCCGAGATGCCAATACATTTAGANAAAATTTACGCGGGTGGCAGCTC	925						
/	119710	GGTGCAAGTGGCTCCCCATTCCCCTGAGAACATAGAGAAGCCCTAAGTAGGGTGGGTACCA	119769:						
/	926	TGGTGGTCTCTGATTTTCAGTTCACCATCTCTATGCTGGCATCAACTTCTCTGCTGGTCAG	985						
/	119770	CCATCGTGTCTGCTTCTCTCACTTTACTCTATGCGGCAATCAACATGTTCTGCTGGTCAG	119829						
/	986	CTTTGCAAGTTGAGTTGGCAGACAGAGATCTCGTCCGCAAGGGCAGAACTGGGGACATA	1045						
/	119830	CTGTTTCAGCTGAAAAATCGACAATCCGGAGCTCATCAGCAAGTCCCGAAGCTGGTACC	119889						
/	1046	TGGGCTGCACTATAGTGTGAGGTTGGTAGAAGTGTGATCATGCTCTGGTTTTAAGT	1105						
/	119890	TGCTCATTTACTACATGACGAGATTATCATGAGAATCCGTCCTCTGCTGTGGTATT	119949						
/	1106	TCCTTTGGAGTGAAGTGTTCATGMAATTAATCTGTCATTTCTTGATTCCTTGAGCTCATT	1165						
/	119950	TTTTTTAAACTGACATATACATGATGTGTGGCCCTCTGTTGATCTTGAGCTGCTTA	120009						
/	1166	TTGCTTATCTGATTTCCATTGACTTCATGCTCTTTCTTCTCCAGTACTTGATTCATTGC	1225						
/	120010	TTGGGTACTGCACCTGGCACTTCTTTCATGCTGTGTTCTATCAGTTTTTCCACCTTGCA	120069						
/	1226	GCTCACTCTTTCACCAATATGT	1247						
/	120070	AAAAGCTCTTCTCCCAAGTGT	120091						

## RESULT 14

[illegible]

AC016752

AC01  
HTC

**KEYWORDS:** HTG. Homo sapiens

**SOURCE ORGANISM**

## DISCUSSION

## REFERENCE

## AUTHORS

[illegible]JOURNAL  
MEDICINE

**MEDLINE  
PURMED**

## REFERENCE

## AUTHORS

**TITLE**

JOURNAL

## REFERENCE.

4

**AUTHORS**  
**TITLE**  
**JOURNAL**

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE	AUTHORS	TITLE	JOURNAL
1	...	...	...
2	...	...	...
3	...	...	...
4	...	...	...
5	...	...	...
6	...	...	...
7	...	...	...
8	...	...	...
9	...	...	...
10	...	...	...
11	...	...	...
12	...	...	...
13	...	...	...
14	...	...	...
15	...	...	...
16	...	...	...
17	...	...	...
18	...	...	...
19	...	...	...
20	...	...	...
21	...	...	...
22	...	...	...
23	...	...	...
24	...	...	...
25	...	...	...
26	...	...	...
27	...	...	...
28	...	...	...
29	...	...	...
30	...	...	...
31	...	...	...
32	...	...	...
33	...	...	...
34	...	...	...
35	...	...	...
36	...	...	...
37	...	...	...
38	...	...	...
39	...	...	...
40	...	...	...
41	...	...	...
42	...	...	...
43	...	...	...
44	...	...	...
45	...	...	...
46	...	...	...
47	...	...	...
48	...	...	...
49	...	...	...
50	...	...	...
51	...	...	...
52	...	...	...
53	...	...	...
54	...	...	...
55	...	...	...
56	...	...	...
57	...	...	...
58	...	...	...
59	...	...	...
60	...	...	...
61	...	...	...
62	...	...	...
63	...	...	...
64	...	...	...
65	...	...	...
66	...	...	...
67	...	...	...
68	...	...	...
69	...	...	...
70	...	...	...
71	...	...	...
72	...	...	...
73	...	...	...
74	...	...	...
75	...	...	...
76	...	...	...
77	...	...	...
78	...	...	...
79	...	...	...
80	...	...	...
81	...	...	...
82	...	...	...
83	...	...	...
84	...	...	...
85	...	...	...
86	...	...	...
87	...	...	...
88	...	...	...
89	...	...	...
90	...	...	...
91	...	...	...
92	...	...	...
93	...	...	...
94	...	...	...
95	...	...	...
96	...	...	...
97	...	...	...
98	...	...	...
99	...	...	...
100	...	...	...

REFERENCE	AUTHORS	TITLE	JOURNAL
1	...	...	...
2	...	...	...
3	...	...	...
4	...	...	...
5	...	...	...
6	...	...	...
7	...	...	...
8	...	...	...
9	...	...	...
10	...	...	...
11	...	...	...
12	...	...	...
13	...	...	...
14	...	...	...
15	...	...	...
16	...	...	...
17	...	...	...
18	...	...	...
19	...	...	...
20	...	...	...
21	...	...	...
22	...	...	...
23	...	...	...
24	...	...	...
25	...	...	...
26	...	...	...
27	...	...	...
28	...	...	...
29	...	...	...
30	...	...	...
31	...	...	...
32	...	...	...
33	...	...	...
34	...	...	...
35	...	...	...
36	...	...	...
37	...	...	...
38	...	...	...
39	...	...	...
40	...	...	...
41	...	...	...
42	...	...	...
43	...	...	...
44	...	...	...
45	...	...	...
46	...	...	...
47	...	...	...
48	...	...	...
49	...	...	...
50	...	...	...
51	...	...	...
52	...	...	...
53	...	...	...
54	...	...	...
55	...	...	...
56	...	...	...
57	...	...	...
58	...	...	...
59	...	...	...
60	...	...	...
61	...	...	...
62	...	...	...
63	...	...	...
64	...	...	...
65	...	...	...
66	...	...	...
67	...	...	...
68	...	...	...
69	...	...	...
70	...	...	...
71	...	...	...
72	...	...	...
73	...	...	...
74	...	...	...
75	...	...	...
76	...	...	...
77	...	...	...
78	...	...	...
79	...	...	...
80	...	...	...
81	...	...	...
82	...	...	...
83	...	...	...
84	...	...	...
85	...	...	...
86	...	...	...
87	...	...	...
88	...	...	...
89	...	...	...
90	...	...	...
91	...	...	...
92	...	...	...
93	...	...	...
94	...	...	...
95	...	...	...
96	...	...	...
97	...	...	...
98	...	...	...
99	...	...	...
100	...	...	...

REFERENCE	AUTHORS	TITLE
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

**COMMENT**

Waterston, R.H.  
Direct Submission  
Submitted (04-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 166436)  
Waterston, R.H.  
Direct Submission  
Submitted (23-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 166436)  
Waterston, R.H.  
Direct Submission  
Submitted (24-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 166436)  
Waterston, R.H.  
Direct Submission  
Submitted (25-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
7 (bases 1 to 166436)  
Waterston, R.  
Direct Submission  
Submitted (30-SEP-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Apr 23, 2000 this sequence version replaced gi:6524293.

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics -----  
Center project name: H NH0506M09

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

**SOURCE INFORMATION:**

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frenken, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-221K4. Actual start of this clone is at base position 1 of RP11-506M9; actual end is at base position 166436 of RP11-506M9.

## FEATURES

Source

Location/Qualifiers

1. 166436  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="Y"  
/map="Y"  
/clone="RP11-506M9"  
/clone.lib="RPC1-11"  
195. 323  
/rpt\_family="L1"  
924. 1077  
/rpt\_family="L1"  
1060. 1830  
/rpt\_family="L1"  
1879. 1918  
/rpt\_family="L1"  
1905. 3257  
/rpt\_family="L1"  
3260. 3393  
/rpt\_family="Alu"  
3618. 3694  
/rpt\_family="MIR"  
4250. 4511  
/rpt\_family="Alu"  
4512. 5804  
/rpt\_family="L1"  
5804. 7682  
/rpt\_family="L1"  
8188. 8294  
/rpt\_family="L2"  
9429. 9499  
/rpt\_family="L2"  
9826. 9872  
/rpt\_family="A)n"  
9962. 10035  
/rpt\_family="L2"  
10596. 10684  
/rpt\_family="CA-rich"  
12836. 13209  
/rpt\_family="MaLR"  
14667. 14689  
/rpt\_family="AT-rich"  
16692. 16712  
/rpt\_family="AT-rich"  
17516. 17584  
/rpt\_family="MERS3"  
17590. 17667  
/rpt\_family="MERS3"  
17814. 18112  
/rpt\_family="Alu"  
19421. 19650  
/rpt\_family="L1"  
19689. 19813  
/rpt\_family="Retroviral"  
19815. 20107  
/rpt\_family="Retroviral"  
20410. 20861  
/rpt\_family="Retroviral"  
21083. 25178  
/rpt\_family="Retroviral"  
25196. 27960  
/rpt\_family="Retroviral"  
28006. 28520  
/rpt\_family="Retroviral"  
28828. 28905  
/rpt\_family="L2"  
29157. 29395  
/rpt\_family="Alu"  
29396. 29438  
/rpt\_family="A)n"  
30965. 30996  
/rpt\_family="AT-rich"  
31776. 32149

repeat\_region  
33205. 33489  
/rpt\_family="Alu"  
35802. 35832  
/rpt\_family="AT-rich"  
36311. 36389  
/rpt\_family="(TTAAA)n"  
36757. 36827  
/rpt\_family="MERS3"  
38147. 38551  
/rpt\_family="Retroviral"  
38552. 38576  
/rpt\_family="(TG)n"  
38577. 38633  
/rpt\_family="Retroviral"  
38634. 39399  
/rpt\_family="Retroviral"  
39400. 39459  
/rpt\_family="MERI\_type"  
39460. 40129  
/rpt\_family="Retroviral"  
40121. 41902  
/rpt\_family="Retroviral"  
41931. 42280  
/rpt\_family="Retroviral"  
42281. 42759  
/rpt\_family="Retroviral"  
42961. 43305  
/rpt\_family="L1"  
43410. 43703  
/rpt\_family="Alu"  
44044. 44848  
/rpt\_family="L1"  
44853. 45060

Query Match 16.08; Score 215.6; DB 9; Length 166436;

Best Local Similarity 62.18; Pred. No. 1.1e-50;  
Matches 390; Conservative 0; Mismatches 234; Indels 4; Gaps 3;

Qy 603 AGTTGTGCTAATGGTATTTCCCTGGTATCTGTCCACCTATGGGGCCACCCCTTTGCAATAT 662  
Db 114404 AGCATTGCTGATGACATGTTCTTCTTACCACTTGTTGATGGGCCATTCACTGCAATAC 114345  
Qy 663 GTTGGCTATCCAGATCAAGTAGATGACTACAAGATTCGCTTGGGGCCACTAGAGTCTT 722  
Db 114344 ACTGGCCATCCATACCAGCAACGACTTTCCTGACATTAAGCTACAAGCGATAAAATTCAT 114285  
Qy 723 CTGCATCACCATCTGGCGGACATGGAGATCACTTCCCGCTCTCTGATTCGTGCTCTT 782  
Db 114284 CTG--TCATGATACGGTGTCTCTTGGTGATTATCTCACGTGTAGTACTCTGGAATTTT 114227  
Qy 783 CTCAGCCACTTTGAAATTTGAAGGCTGTGCCCTTCTAGTGTCTCAACTTCTCTGATCATCT 842  
Db 114226 TCCTGCACTCTGAAACAGAGAGCCCTACCTTACTATTATCATATATTTTGTATTAT 114167  
Qy 843 CTTTGAGCCCTGGATTAAAGTTCTGGAGAAAGTGGTGGCCAGATGCCCAATAACATTGAGAA 902  
Db 114166 GTTGGCACCATGGCTGGAGTTTGGAAAAGTGGAGCTCATCTTCAACAACACAGAAAA 114107  
Qy 903 AAATTCAGCCGGTCCGCACTCTGGTGGTCTCTGATTTTCAGTCACCACTCTCTATGCTGG 962  
Db 114106 TAATTCAGCATGTGGGTAC-ATACGGATGCTTACCTTAATCTACTATATGCTGTC 114048  
Qy 963 CATCAACTTCTTTGCTGGTCCAGCTTTTCAGTTGAGTTGGCAGACAGAGATCTCGTCA 1022  
Db 114047 TATCAACTTCTCTGGTGGTCCAGCACTGAAACTGCAGCTCTCAATAGGAGATTAATGA 113988  
Qy 1023 CAAGGGCAGAACTGGGACATATGGGCTGCATATAGTGTGAGTTGGTGAAGATGT 1082  
Db 113987 CAAGAGACTGAGTGGGC-CATAGAATCCTACACTACAGCTCTTAGTTTTTAGAAAAATGT 113929  
Qy 1083 GATCATGCTCTGGTTTTTAAGTTCTTTGGAGTGAAGTTTACTGAATTTACTGTCAATC 1142

113928 GATAATAATGATATTATGTTGTTGGAGGAAAAATTTCACTGAAGTCTTGACTC 113869  
 1143 CTTGATTGCTTCGAGCTCAATTATTGTTATCTGATTTCATTGCTCATGCTCCCTTTT 1202  
 113868 ATTAATGCGCATGATGTTATCATCACTACTACTATTAGCCATTGGCTTTATGCTCCCTTT 113809  
 1203 CTTCCAGTACTTCATCCATTCGCGCTCA 1230  
 CTGTCAGTATTGCACCAAGTGGTCA 113781

SULT 15  
 AC017005 161879 bp DNA linear PRI 09-MAY-2001  
 AC017005 Homo sapiens BAC clone RP11-100J21 from Y, complete sequence.  
 AC017005 AC017005.6 GI:13435269  
 AC017005 HTG.  
 AC017005 YWORDS  
 AC017005 YORCE  
 AC017005 ORGANISM  
 AC017005 Homo sapiens.  
 AC017005 Homo sapiens  
 AC017005 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AC017005 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AC017005 Sulston,J.E. and Waterston,R.  
 AC017005 1 (bases 1 to 161879)  
 AC017005 Toward a complete human genome sequence  
 AC017005 Genome Res. 8 (11), 1097-1108 (1998)  
 AC017005 99063792  
 AC017005 PUBMED  
 AC017005 2 (bases 1 to 161879)  
 AC017005 Armstrong,J., Stonking,T., Hawkins,M. and Bernard,K.  
 AC017005 The sequence of Homo sapiens BAC clone RP11-100J21  
 AC017005 Unpublished  
 AC017005 3 (bases 1 to 161879)  
 AC017005 Waterston,R.H.  
 AC017005 Direct Submission  
 AC017005 Submitted (09-DEC-1999) Genome Sequencing Center, Washington  
 AC017005 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 AC017005 MO 63108, USA  
 AC017005 4 (bases 1 to 161879)  
 AC017005 Waterston,R.H.  
 AC017005 Direct Submission  
 AC017005 Submitted (23-MAR-2001) Genome Sequencing Center, Washington  
 AC017005 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 AC017005 MO 63108, USA  
 AC017005 5 (bases 1 to 161879)  
 AC017005 Waterston,R.  
 AC017005 Direct Submission  
 AC017005 Submitted (09-MAY-2001) Department of Genetics, Washington  
 AC017005 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 AC017005 On Mar 23, 2001 this sequence version replaced gi:9211335.  
 AC017005 ----- Genome Center  
 AC017005 Center: Washington University Genome Sequencing Center  
 AC017005 Center code: WUGSC  
 AC017005 Web site: http://genome.wustl.edu/gsc  
 AC017005 Contact: sapiens@watson.wustl.edu  
 AC017005 ----- Summary Statistics  
 AC017005 -----  
 AC017005 Center project name: H\_NH0100J21  
 AC017005 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)

VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-79J10; the clone sequenced to the right is RP11-245K4, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-100J21; actual end is at base position 66964 of RP11-245K4.

#### FEATURES

source  
 1. .161879  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="Y"  
 /map="Y"  
 /clone="RP11-100J21"  
 /clone\_lib="RPCI-11"  
 1. .1185  
 /rpt\_family="L1"  
 1183. .1513  
 /rpt\_family="L1"  
 1679. .1842  
 /rpt\_family="MER1\_type"  
 2125. .3135  
 /rpt\_family="ERV1"  
 3116. .3412  
 /rpt\_family="ERV1"  
 3450. .3866  
 /rpt\_family="ERV1"  
 3867. .6977  
 /rpt\_family="ERV1"  
 6986. .7295  
 /rpt\_family="ERV1"  
 7297. .7625  
 /rpt\_family="ERV1"  
 9507. .9570  
 /rpt\_family="MER1\_type"  
 11143. .11817  
 /rpt\_family="L1"  
 12324. .12843  
 /rpt\_family="L1"  
 12874. .12922  
 /rpt\_family="ERV1"  
 12885. .12929  
 /rpt\_family="ERV1"  
 13051. .13264  
 /rpt\_family="L1"  
 13381. .13619  
 /rpt\_family="L1"  
 13628. .14306  
 /rpt\_family="L1"  
 15532. .15662  
 /rpt\_family="ERV1"  
 15740. .16044  
 /rpt\_family="ERV1"  
 16045. .16136  
 /rpt\_family="ERV1"  
 16178. .16239  
 /rpt\_family="ERV1"  
 16728. .17201

BEST AVAILABLE COPY

```
repeat_region /rpt_family="ERV1"
17188..17253
repeat_region /rpt_family="ERV1"
17243..17602
repeat_region /rpt_family="ERV1"
17746..18064
repeat_region /rpt_family="L1"
18085..18235
repeat_region /rpt_family="L1"
18303..18876
repeat_region /rpt_family="ERVK"
18927..19413
repeat_region /rpt_family="ERVK"
19419..19628
repeat_region /rpt_family="L1"
21508..21759
repeat_region /rpt_family="Alu"
22725..22868
repeat_region /rpt_family="ERV1"
22868..23335
repeat_region /rpt_family="ERV1"
23350..24076
repeat_region /rpt_family="ERV1"
24273..24690
repeat_region /rpt_family="ERV1"
24702..24823
repeat_region /rpt_family="ERV1"
24886..25175
repeat_region /rpt_family="Alu"
25186..27543
repeat_region /rpt_family="ERV1"
27654..27697
repeat_region /rpt_family="ERV1"
28186..28421
repeat_region /rpt_family="ERV1"
28440..28804
repeat_region /rpt_family="ERV1"
28773..29267
repeat_region /rpt_family="ERV1"
29270..31939
repeat_region /rpt_family="ERV1"
31933..32211
repeat_region /rpt_family="ERV1"
32225..32546
repeat_region /rpt_family="Alu"
32582..33624
repeat_region /rpt_family="ERV1"
33740..34049
repeat_region /rpt_family="ERV1"
34050..34149
repeat_region /rpt_family="ERV1"
34155..34617
repeat_region /rpt_family="ERV1"
34624..37556
repeat_region /rpt_family="ERV1"
37676..37979
repeat_region /rpt_family="ERV1"
38072..38379
repeat_region /rpt_family="ERV1"
38380..39312
repeat_region /rpt_family="ERV1"
39318..40245
repeat_region /rpt_family="ERV1"
40372..40443
repeat_region /rpt_family="ERV1"
40486..41009
repeat_region /rpt_family="ERV1"
```

Query Match 15.9%; Score 214; DB 9; Length 161879;  
Best Local Similarity 61.9%; Pred. No. 3.3e-50;  
Matches 389; Conservative 0; Mismatches 235; Indels 4; Gaps 3;  
QY 603 AGTTGCTAATGTTATTTCCCTGGTATCTGTCCACCTATGGGGCCACCCTTTGCAATAT 662

```
Db 160312 AGATTGCTGATGACATGTTCTCTTTACCACTTGTATGCGGCCATTCACTGCAATAC 160371
QY 663 GTTGGCTATCCAGATCAAGTAGCATGACTACAAGATTTCGGCTTGGGCCACTAGAAATCCT 722
Db 160372 ACTGGCCATCCATACCAGCAACGACTTTCCTGACATTTAGCTACAAGCGATAAATTCAT 160431
QY 723 CTGATCAACATCTCGCGGACATTTGAGATCACTTCCCGCTCTCTGAATTCGTGCTCTT 782
Db 160432 CTG--TCATGATACGGTGTCTCTTGGTATTTATCTCACGTGTAGTACTCTGGAATTTT 160489
QY 783 CTCAGCCATTTGAAATTTGAAGCTGTGCCCTTCCCTAGTCTCAACTTCCCTGATCATCCT 842
Db 160490 TCTGTCATCTCTGAAACAGAGAGCCTACCAATTAATTAATTAATTAATTTTGTATTAT 160549
QY 843 CTTTGAGCCCTGGATTAAGTTCTGAGAAAGTGTGGCCAGATGCCCAATTAACATTTGAGAA 902
Db 160550 GTTGGCACCATGGCTGGAGTTTTTGGAAAGTGGAGCTCATCTTCATAACAACAGAGAAA 160609
QY 903 AAACITTCAGCGGGTCCGGCACTCTGTGTCTGTGTTTTCAGTCACCATCTCTATGCTGG 962
Db 160610 TAATTCAGCATGGTGGGTAC-ATACGGATGCTTACCTTAATCATGCTACTATATGCTGC 160668
QY 963 CATCACTTCTCTTGTGTGCTGAGCTTTGAGTTGAGTTGGCAGACAGAGATCTCGTCA 1022
Db 160669 TATCAACTTCTCTGGCTGGTCAAGCTGAAACTGCAGCTCTCAATGAGGAAGTAATGA 160728
QY 1023 CAAAGGGCAGAACTGGGGACATATGGGCCCTGCACATAGTGTGAGGTTGGTAGAGAAATGT 1082
Db 160729 CAAGAGACTGAGGTGGGC-CATAGAATCCTACACTACAGCTTCTAGTTTTTAGAAAATGT 160787
QY 1083 GATCATGGTCTTGGTTTTTAAAGTTCTTTGGAGTGAAGTGTACTGAATTAAGTGTCTTC 1142
Db 160788 GATAATAATATTGATATTTATGTTTGGTGGAGGAAAAATTCACAGAAGTCTTGTGACTC 160847
QY 1143 CTTGATTGCTTGCAGCTCATTTATTTGCTTATCTGATTTCCATTGACCTCATGCTCCTTT 1202
Db 160848 ATTAATTGCCATGTAGTTCATCAATAACCTTAACCTATTAGCCATTGGCTTTATGCTCTCT 160907
QY 1203 CTTCCAGTACTTGCATCCATTGGGCTCA 1230
Db 160908 CTGTCAGTATTTGCACCCCAAGTGGTCA 160935
```

Search completed: March 30, 2003, 05:11:04  
Job time : 3911.69 secs